



Avaliação da Precisão dos Métodos ARIMA, Suavização Exponencial e Redes Neurais na Previsão de Séries Temporais Anuais Brasileiras

Discente: Gustavo Inocência Camilo

Orientador: Prof. Dr. Marcelo Ruy

Resumo:

Dentre as aplicações de Análise de Séries Temporais, a previsão de valores futuros é uma das mais utilizadas. Nas previsões, a precisão do método utilizado é um dos fatores críticos em sua adoção. Três dos principais métodos de previsão disponíveis, devido ao uso generalizado e à qualidade das previsões, são a suavização exponencial, o modelo ARIMA e as redes neurais artificiais. Este artigo tem como objetivo testar a hipótese de igualdade ou não entre as precisões das previsões pontuais geradas por estes três métodos quando aplicados a séries temporais anuais brasileiras contemporâneas. Para tanto, 40 séries anuais do IPEA foram selecionadas por meio de amostragem aleatória simples, previsões com os três métodos foram executadas e suas precisões foram calculadas. A conclusão foi que os três métodos têm em média o mesmo grau de precisão.

Palavras chave: Séries Temporais Univariadas; Métodos de Previsão; Comparação de Métodos; Erros de Previsão.

Evaluating Accuracy of the Methods ARIMA, Exponential Smoothing and Neural Networks in Forecasting Brazilian Annual Time Series

Abstract:

Among the applications of Time Series Analysis, forecasts of future values is one of the most used. In forecasting, the precision of the method used is one of the critical factors in its adoption. Three of the major forecasting methods available, due to the widespread use and quality of predictions, are exponential smoothing, ARIMA model, and artificial neural networks. This article aims to test the hypothesis of equality or not between the precisions of point forecasts generated by these three methods when applied to contemporaneous Brazilian time series. For this purpose, 40 annual series of IPEA were selected by means of simple random sampling, predictions with the three methods were executed and their precisions were calculated. The conclusion was that the three methods have the same degree of accuracy on average.

Key words: Univariate Time Series; Forecast Methods; Comparison of Methods; Forecast Errors.

1. Introdução

Segundo Morettin e Tolo (2006), uma série temporal é qualquer variável cujas observações foram coletadas sequencialmente ao longo do tempo. Quando o registro é feito continuamente, a série é dita contínua. Caso as observações sejam tomadas em pontos específicos do tempo, normalmente igualmente espaçados, a série é denominada discreta. Isto é, contínua e discreta se relacionam ao registro das observações e não à natureza da variável sendo medida. Os métodos estudados neste artigo são voltados a séries temporais discretas tomadas em intervalos regulares de tempo.

Os principais objetivos da análise de séries temporais são: descrever o comportamento da série, investigar seu mecanismo gerador, procurar periodicidades nos dados e fazer previsões de seus



valores futuros (MORETTIN; TOLOI, 2006). Na prática, este último tópico é um dos mais utilizados, principalmente na área de Gestão de Operações, com especial ênfase nas aplicações de previsão da demanda. Neste contexto, a precisão das previsões é um fator crítico para reduzir custos e fornecer um melhor serviço aos clientes (MAKRIDAKIS; HIBON, 2000).

De acordo com Hyndman e Athanasopoulos (2018) os principais métodos estatísticos utilizados para a previsão de séries temporais são a suavização exponencial e o modelo autorregressivo integrado de médias móveis (ARIMA – *autoregressive integrated moving average*).

Previsões feitas pelos métodos de suavização exponencial são médias ponderadas de valores passados, com os pesos decaindo exponencialmente à medida que as observações se tornam mais antigas. Já o modelo ARIMA explora outra característica distintiva das séries temporais: a presença de autocorrelação serial, que seria o grau de associação linear entre valores defasados da mesma. O modelo ARIMA utiliza esta autocorrelação em diversas defasagens para prever os valores futuros da série.

Mais recentemente, vem ganhando popularidade métodos baseados em *machine learning*, principalmente aqueles que utilizam redes neurais artificiais (RNA). Isto pode ser verificado comparando-se a quantidade destes métodos na *M4 Competition* (MAKRIDAKIS; SPILLOTIS; ASSIMAKOPOULOS, 2018) com sua antecessora, a *M3 Competition* (MAKRIDAKIS; HIBON, 2000). Estas foram competições onde *experts* foram convidados a fornecer previsões para séries temporais reais usando métodos variados.

Um ponto negativo para a adoção prática destes três tipos de métodos sempre foi o seu grau de complexidade teórica. Mas, com o atual aumento do poder computacional, aliado à crescente disponibilidade de *software* estatístico dedicado à análise de séries temporais, este panorama está mudando. Já é possível encontrar diversas soluções completamente automatizadas; basta ao usuário fornecer a série temporal e o método desejado que o programa sozinho seleciona o modelo mais adequado, estima os seus parâmetros e executa as previsões.

Dentro desta categoria de soluções automatizadas, destaca-se o pacote *forecast* de Hyndman e Khandakar (2008), escrito na linguagem de programação R (R CORE TEAM, 2019). Pacotes são coleções de funções, dados e códigos compilados que ampliam as capacidades originais do *software* R. Normalmente são programados por estatísticos computacionais e não raro são o estado da arte em determinada área da Estatística. Como os pacotes e o próprio *software* são de código aberto, eles estão disponíveis ao escrutínio da comunidade científica com relação a sua correção, eficácia, etc. O pacote *forecast* tem uma grande quantidade de funções automatizadas voltadas para a análise e previsão de séries temporais, incluindo a suavização exponencial, o modelo ARIMA e as RNA, o que nos remete ao próximo tópico.

2. Objetivo

Este trabalho tem como objetivo testar se há ou não diferença significativa entre as precisões das previsões pontuais geradas pelos métodos automáticos de suavização exponencial, ARIMA e RNA, quando aplicados a séries temporais anuais brasileiras contemporâneas.

Muito embora haja trabalhos que comparem diversos métodos de previsão tais como Makridakis e Hibon (2000) e Makridakis, Spiliotis e Assimakopoulos (2018), os mesmos utilizam séries estrangeiras. Devido às muitas particularidades do Brasil (econômica, social, demográfica, etc.), seria interessante estudar o desempenho dos métodos em nosso contexto em particular. Como a pesquisa é de caráter aplicado, o foco é em dados recentes (coletados a partir de 1990). Além disso, uma vez que as séries são anuais, não há sazonalidade presente.



3. Referencial Teórico

Nesta seção serão abordados os métodos quantitativos de previsão comparados neste trabalho, a saber, suavização exponencial, modelo ARIMA e redes neurais artificiais, bem como medir a precisão das previsões executadas pelos métodos.

3.1. Métodos Quantitativos de Previsão

Segundo Hyndman e Athanasopoulos (2018), o método de previsão mais apropriado depende principalmente do tipo de dado disponível. Quando não há dados ou se os existentes não são relevantes para a previsão em questão (por exemplo, a demanda por um produto totalmente novo), métodos qualitativos são mais indicados. Tais métodos são estruturados para a obtenção de previsões sem o uso de dados históricos.

Os métodos quantitativos podem ser aplicados quando duas condições são satisfeitas: (i) informações numéricas a respeito do passado estão disponíveis; (ii) é razoável supor que alguns aspectos dos padrões passados continuarão no futuro. Três dos principais métodos de previsão quantitativos disponíveis, devido ao uso generalizado e à qualidade das previsões, são a suavização exponencial, o modelo ARIMA e as redes neurais artificiais (HYNDMAN; ATHANASOPOULOS, 2018).

De acordo com estes autores, os métodos de suavização exponencial decompõem a série em componentes e modelam cada componente usando relações recursivas. Os componentes ou padrões de uma série temporal são tendência, ciclos, sazonalidade e variações irregulares. A tendência pode ser definida como uma mudança de longo prazo na média da série. A sazonalidade é um padrão que se repete com uma periodicidade conhecida e fixa, normalmente anual. Nesses casos, séries com frequência anual ou superior não possuem sazonalidade. Um ciclo é um padrão que se repete com alguma regularidade, mas com uma periodicidade não fixa, normalmente longa e às vezes até desconhecida. Tradicionalmente, inclusive na suavização exponencial, o componente cíclico é agrupado à tendência e ambos são denominados genericamente de tendência. Por fim, as variações irregulares ou erro é o que resta após a eliminação dos outros fatores.

Os métodos mais conhecidos de suavização exponencial são a suavização exponencial simples, a suavização de Holt e a suavização de Holt-Winters. A primeira é usada em séries localmente constantes (sem tendência e sem sazonalidade), a segunda em séries com tendência apenas e a última em séries com sazonalidade, sendo esta aditiva (para séries com amplitude constante) ou multiplicativa (quando a amplitude varia ao longo do tempo).

De acordo com Hyndman e Athanasopoulos (2018), estes métodos de suavização assim como propostos originalmente por Brown, Holt e Winters no final da década de 1950 não tem base probabilística, sendo apenas relações recursivas e por isso não geram, por exemplo, intervalos de predição, somente estimativas pontuais. Atualmente, após o trabalho de diversos autores, a suavização exponencial foi expandida e ganhou uma base probabilística denominada modelo de espaço de estados. Hyndman et al. (2008) apresentam uma classificação dos métodos de suavização: combinando o erro com a tendência e a sazonalidade existem 30 modelos de suavização, onde os métodos de suavização simples, de Holt e de Holt-Winters são casos especiais. Tal classificação é denominada pelos autores de ETS (*error, trend and seasonality*) e é implementada no pacote *forecast*. O próprio pacote determina automaticamente qual dos 30 modelos é o que melhor se ajusta aos dados, minimizando uma função perda. O quadro 1 mostra a classificação proposta por Hyndman et al. (2008).

Para cada um dos 15 métodos do quadro 1 existem duas variantes, uma com o erro aditivo e outra com o erro multiplicativo, gerando 30 métodos de suavização. A suavização exponencial

simples é o modelo (N,N), a suavização de Holt, o modelo (A,N) e a de Holt-Winters, os modelos (A,A) e (A,M).

Quadro 1: Classificação dos Métodos de Suavização Exponencial

Tendência	Sazonalidade		
	N	A	M
N (nenhuma)	N,N	N,A	N,M
A (aditiva)	A,N	A,A	A,M
Ad (aditiva e amortecida)	Ad,N	Ad,A	Ad,M
M (multiplicativa)	M,N	M,A	M,M
Md (multiplicativa e amortecida)	Md,N	Md,A	Md,M

Fonte: Hyndman et al. (2008), p. 12.

De acordo com Hyndman e Athanasopoulos (2018), ao contrário do método anterior que modela os componentes da série, o modelo ARIMA explora outra característica distintiva das mesmas: o fato de dados próximos terem maior relação entre si do que dados separados por grandes intervalos de tempo. Esta característica é resumida matematicamente por uma grandeza denominada autocorrelação serial, que seria o grau de associação linear entre valores defasados da série. A autocorrelação na defasagem 1 (dados separados por 1 unidade de tempo) seria a associação entre y_t e y_{t-1} , na defasagem 2 entre y_t e y_{t-2} e assim sucessivamente. O modelo ARIMA utiliza esta autocorrelação em diversas defasagens para prever os valores futuros da série.

Como o presente artigo tem como foco séries sem sazonalidade, será apresentado apenas o modelo ARIMA não sazonal. A seguir são explicados os componentes do modelo ARIMA: os termos autoregressivos (AR), os termos de médias móveis (MA) e o nível de integração da série (I).

Segundo Hyndman e Athanasopoulos (2018), em um modelo autorregressivo, a variável de interesse (y_t) é escrita como uma combinação linear de seus valores passados (y_{t-1} , etc.). Um modelo autorregressivo de ordem p é representado por $AR(p)$. Ao invés de utilizar os valores passados da variável de interesse, um outro tipo de modelo utiliza uma média ponderada de distúrbios aleatórios presente e passados na equação (ε_t , ε_{t-1} , etc.). Este modelo, é denominado de médias móveis de ordem q e é representado por $MA(q)$. A combinação do modelo autorregressivo $AR(p)$ com o de médias móveis $MA(q)$ resulta no modelo denominado $ARMA(p,q)$. Para muitas séries encontradas na prática, se quisermos um modelo com um número não muito grande de parâmetros, a inclusão de termos autorregressivos e de médias móveis é uma solução adequada. Os modelos $AR(p)$ e $MA(q)$ são casos especiais do modelo $ARMA(p,q)$ tomando-se $q = 0$ e $p = 0$, respectivamente.

Os 3 modelos apresentados anteriormente são para séries estacionárias na média (i.e., com média constante). Séries com tendência estocástica podem ser transformadas em séries estacionárias na média tomando-se diferenças entre seus valores sucessivos. Se após aplicarmos a 1ª diferença, a série estabilizar sua média, ela é dita integrada de ordem 1. Caso isso não ocorra, aplica-se a 2ª diferença e assim sucessivamente até que se obtenha uma série com média constante. Uma série integrada de ordem d é aquela que necessitou de d diferenças para tornar-se estacionária na média. Se uma série integrada de ordem d puder ser modelada por um processo $ARMA(p,q)$, teremos um modelo $ARIMA(p,d,q)$ não sazonal, cuja equação é dada por:

$$\Delta^d y_t = c + \phi_1 \cdot \Delta^d y_{t-1} + \dots + \phi_p \cdot \Delta^d y_{t-p} + \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1} + \dots + \theta_q \cdot \varepsilon_{t-q} \quad (1)$$



Na equação (1), y é a variável de interesse, c é o intercepto, Δ^d é a quantidade de diferenciações (se não forem necessárias, são excluídas da equação), ε é uma sequência de variáveis aleatórias independentes igualmente distribuídas denominada de erro e os termos ϕ e θ são parâmetros a serem estimados a partir dos dados.

O pacote *forecast* determina automaticamente se diferenciações são necessárias utilizando o chamado de teste de raízes unitárias. A seguir, o mesmo seleciona as ordens p e q minimizando uma função perda, por padrão o critério de informação de Akaike corrigido (AICc). Por fim, os parâmetros ϕ e θ são estimados pelo método de máxima verossimilhança. (HYNDMAN; KHANDAKAR, 2008).

Hyndman e Athanasopoulos (2018) explicam que Redes Neurais Artificiais (RNA) são métodos baseados em modelos matemáticos simplificados do cérebro. Elas permitem modelar relações complexas e não lineares entre a resposta (saída) e seus previsores (entradas). Uma RNA pode ser considerada uma rede de “neurônios” (nós) organizada em camadas. As entradas são a camada inferior e a saída a superior. Pode haver, também, camadas intermediárias contendo neurônios ocultos.

A rede mais simples não contém camadas ocultas e é equivalente à regressão linear. Os coeficientes relativos a cada entrada são denominados pesos e a saída é obtida combinando-se linearmente as entradas. Os pesos são determinados utilizando um algoritmo de aprendizagem que minimiza uma função perda (p. ex. erro quadrático médio). Uma vez introduzida uma ou mais camadas intermediárias com neurônios ocultos, a rede se torna não linear. Isto é conhecido como rede com propagação progressiva de múltiplas camadas, onde uma camada é a entrada para a próxima. As entradas em cada nó são calculadas usando uma combinação linear e o resultado é então modificado por uma função não linear (p. ex., função sigmoide).

Os pesos iniciais são aleatórios e são atualizados utilizando os dados de treino. Assim, uma rede é normalmente treinada diversas vezes usando diferentes valores iniciais e tira-se uma média dos resultados obtidos. O número de camadas ocultas e de nós em cada camada devem ser especificados a priori.

Na análise de séries temporais, os valores defasados da série podem ser usados como entradas da RNA. Isto é denominado modelo de rede neural autoregressiva. O pacote *forecast* implementa este modelo utilizando uma única camada oculta e, para dados não sazonais, p valores defasados da série. Quando não especificado (automático), o valor de p é determinado minimizando o critério de informação de Akaike (AIC) e quantidade de nós é o inteiro mais próximo de $p/2$. Este modelo é equivalente a um $AR(p)$, mas com funções não lineares.

Nas previsões, a rede é aplicada iterativamente. Para as previsões um período a frente, utilizam-se simplesmente os dados históricos como entrada. Para a previsão 2 períodos a frente, utilizam-se a previsão 1 período a frente juntamente com os dados históricos como entrada, e assim sucessivamente até a obtenção de todas as previsões requeridas.

3.2. Avaliação da Precisão das Previsões

Para Hyndman e Athanasopoulos (2018), é importante avaliar a precisão das previsões utilizando previsões genuínas. Dessa forma, os resíduos do modelo não são uma indicação confiável dos erros de previsão. A precisão somente pode ser determinada considerando-se o desempenho do modelo em dados novos e que não foram utilizados em seu ajuste.

Uma solução é separar os dados disponíveis em duas partes: os dados de treino e os dados de teste. O primeiro é usado para estimar os parâmetros do modelo, enquanto o segundo é utilizado para se avaliar a precisão das previsões. Como os dados de teste não foram usados no ajuste do

modelo, eles são um indicador da qualidade do modelo de previsão em dados novos.

O erro de previsão é a diferença entre o valor observado e o previsto. Segundo Hyndman e Athanasopoulos (2018), a precisão das previsões pode ser avaliada de três maneiras: por meio dos erros dependentes de escala, dos erros percentuais ou dos erros escalonados.

Os erros dependentes de escala estão na mesma unidade de medidas da série original. Muito embora sejam mais simples e intuitivos, a desvantagem é que não se pode comparar erros advindos de séries que estejam em unidades diferentes. Os erros mais comuns deste tipo são o erro absoluto médio e a raiz quadrada do erro quadrático médio.

Os erros percentuais não possuem a desvantagem anterior, pois são livres de unidade. Entretanto, são indefinidos quando $y_t = 0$ e só são aplicáveis em séries medidas em uma escala de razão. Os principais são o erro percentual absoluto médio e o erro percentual absoluto médio simétrico.

Uma alternativa aos erros percentuais, por não possuírem as desvantagens anteriores, são os erros escalonados. Dentre os erros escalonados, Hyndman e Athanasopoulos (2018) propõem o uso do erro escalonado absoluto médio (MASE – *mean absolute scaled error*), definido como a média aritmética dos valores absolutos dos q_j dados pela equação 2. O pacote *forecast* calcula os erros aqui citados, incluindo o MASE. Neste artigo, este erro será utilizado para se comparar a precisão dos três métodos de previsão.

$$q_j = \frac{e_j}{\frac{1}{T} \cdot \sum_{t=2}^T |y_t - y_{t-1}|} \quad (2)$$

4. Metodologia

Inicialmente, foram obtidas as séries temporais anuais a serem analisadas. As séries foram retiradas do banco de dados do IPEA – Instituto de Pesquisa Econômica Aplicada (<http://www.ipeadata.gov.br>).

Neste banco de dados, há 8742 séries temporais divididas de acordo com o tema (macroeconômicas, regionais e sociais) e com sua frequência (anual, trimestral, mensal e outras). Destas, 3300 são anuais. Como o objetivo são séries contemporâneas, foram retidas apenas as séries anuais iniciadas a partir de 1990 e que possuíam no mínimo 20 termos, de forma a gerar um conjunto de teste com 6 dados e um conjunto de treino com no mínimo 14 termos. Este foi o procedimento seguido por Makridakis e Hibon (2000) na *M3 Competition*, pois, segundo os autores, 14 é suficiente para estimar os parâmetros dos modelos. Isto gerou um total de 561 séries (484 macroeconômicas, 17 regionais e 60 sociais).

Deste universo de 561, 40 delas foram sorteadas por meio de amostragem aleatória simples para serem analisadas. Uma amostra de tamanho 40 foi escolhida por ser considerada uma amostra de tamanho suficiente para justificar as aproximações de grandes amostras para os testes paramétricos e não paramétricos (MONTGOMERY; RUNGER, 2012).

De posse das 40 séries, cada uma foi dividida em duas partes: dados de treino e de teste. Os dados de teste foram sempre os últimos 6 valores das séries, ou seja, o horizonte de tempo para as previsões foi de 6 anos. A cada uma das 40 séries de treino foram ajustados três modelos: ETS, ARIMA e RNA. Os parâmetros de todos os modelos foram determinados automaticamente pelo pacote *forecast* versão 8.7 conjuntamente com o *software* R versão 3.6.0.

Uma vez ajustados os três modelos aos dados de teste, o *software* fez as previsões 6 períodos a



frente. Dadas as previsões e os dados de teste, o pacote *forecast* calcula diversos erros de previsão, inclusive o MASE utilizado neste trabalho. No Apêndice 1 tem-se os comandos utilizados exemplificados em uma das 40 séries analisadas.

Por fim, de forma a atingir o objetivo do trabalho, estaremos testando as hipóteses abaixo por meio de métodos estatísticos adequados descritos na próxima seção:

H_0 : Os três métodos possuem em média a mesma precisão

H_1 : Pelo menos um método possui em média uma precisão diferente

5. Resultados

A tabela 1 apresenta o MASE calculado para cada uma das 40 séries utilizando os 3 métodos.

Tabela 1: MASE das Previsões dos 3 Métodos

Série	ARIMA	ETS	RNA
1	11,0949	11,0953	10,1437
2	2,7424	2,5380	0,7476
3	3,0706	3,0706	1,2487
4	5,1564	5,1147	4,5682
5	1,8613	1,9067	2,8691
6	7,7044	5,5231	3,0958
7	6,4355	2,6401	2,2343
8	1,6480	1,6480	1,6496
9	9,9557	9,8113	2,0775
10	2,0895	2,8942	2,8571
11	13,9343	2,1313	10,5621
12	2,1779	2,1779	3,4192
13	21,5127	19,7031	14,2635
14	1,4728	0,8022	2,0036
15	0,4146	0,4146	2,2421
16	0,6618	0,6618	0,6909
17	2,2701	0,7867	0,6708
18	1,6327	1,6325	0,6508
19	7,4437	11,1216	3,1823
20	2,0673	2,1229	2,1120
21	6,8283	2,6752	2,3511
22	16,4998	4,9576	22,9179
23	5,9346	1,9640	1,4742
24	2,7947	5,6231	1,8356
25	9,6318	2,8455	9,7054
26	1,2080	1,2080	1,2497
27	1,1372	1,1381	1,2672
28	0,4838	1,3753	1,7157
29	3,8594	3,4605	5,3129
30	10,9248	1,9278	12,6105
31	5,3554	5,8873	4,9371
32	4,0817	4,0816	4,3704
33	1,7430	1,7489	2,9638
34	7,4972	7,5135	7,3881
35	5,5171	7,9631	6,3892
36	3,0108	5,8273	7,3043
37	1,1897	5,6117	6,4562
38	1,6700	5,8076	6,3616
39	2,4047	2,5467	3,2117
40	1,9227	2,0177	2,3337

Fonte: Elaboração Própria



De acordo com Montgomery e Runger (2012), o método paramétrico para se testar a hipótese de que os 3 métodos têm em média a mesma precisão *versus* que as precisões médias são diferentes para pelo menos um deles é a Análise de Variância (ANOVA) com blocos. Como cada série foi analisada por três métodos diferentes, tem-se três amostras dependentes (emparelhadas). O *software* R base tem as funções necessárias para se executar a ANOVA com blocos.

Aplicou-se a ANOVA com blocos nos dados da Tabela 1. Entretanto, antes de se aceitar os resultados é necessário analisar os resíduos da ANOVA e checar por violações nas suposições básicas. Especificamente, as suposições de homoscedasticidade e normalidade dos resíduos não se sustentam. Logo o teste paramétrico não pode ser utilizado, devendo ser substituído por um teste não paramétrico equivalente. Os testes não paramétricos são livres de distribuição, no sentido que seu desenvolvimento não é baseado na hipótese de normalidade (MONTGOMERY; RUNGER, 2012).

De acordo com Siegel e Castellan Jr. (2006), o teste não paramétrico equivalente à ANOVA com blocos é o teste de Friedman. Este teste utiliza os postos (ranques) dos dados ao invés de seus valores brutos. Para se calcular a estatística de teste de Friedman, ordenam-se as observações da menor para a maior de forma separada em cada um dos blocos (linhas da tabela 1) e atribuem-se os ranques {1, 2, 3} para cada bloco. Na sequência, os postos médios por tratamento (cada coluna da Tabela 1) são calculados e comparados. Se todos os tratamentos tiverem o mesmo efeito no nível da resposta (isto é, se H_0 for verdadeira), para cada tratamento haverá uma mistura de postos, de forma que os postos médios por tratamento serão próximos. Por outro lado, se um ou mais tratamentos forem diferente dos demais, eles terão postos mais baixos ou altos de maneira mais ou menos uniforme, de forma que seus postos médios apresentarão uma diferença significativa dos demais, levando à rejeição da hipótese nula.

O *software* R possui uma função que implementa o procedimento anteriormente descrito. Ou seja, fornecidos os dados da Tabela 1, o *software* determina os postos por bloco, calcula a estatística de teste e fornece seu valor-p, que então é utilizado para se julgar a rejeição ou não de H_0 . Aplicando-se este teste aos dados da Tabela 1 encontrou-se um valor-p de 0,6318, superior ao maior nível de significância tradicionalmente utilizado de 10%. Logo, falha-se em rejeitar a hipótese nula H_0 . Ou seja, não há evidências que os três métodos não possuam em média a mesma precisão.

6. Conclusão

Este artigo teve como objetivo testar a hipótese de igualdade ou não entre as previsões das previsões pontuais geradas pelos métodos automáticos ETS, ARIMA e RNA, quando aplicados a séries temporais anuais brasileiras contemporâneas. Para tanto, 40 séries anuais foram selecionadas por meio de amostragem aleatória simples, os três métodos foram ajustados aos dados de treino, previsões com os modelos foram executadas e suas precisões foram calculadas comparando-as com os dados de teste. Utilizando o teste não paramétrico de Friedman, determinou-se que a hipótese de igualdade entre as precisões não pode ser rejeitada. Este resultado requer ser analisado do ponto de vista de seu grau de generalização e de suas limitações.

Do ponto de vista estatístico, o resultado é generalizável para a população de séries anuais das quais a amostra foi retirada, ou seja, as 561 séries anuais do IPEA.

Porém, mais interessante é a generalização do estudo para a teoria. Os resultados do presente trabalho são semelhantes aos observados na *M3* e na *M4 Competition*. Na *M3 Competition*, o desempenho (medido pelo erro percentual absoluto médio simétrico) dos modelos ARIMA foi



muito próximo ao da suavização expoencial (média de 23,9 *versus* 23,7, respectivamente). Já na *M4 Competition*, não houve métodos “puros”, apenas híbridos (combinação de dois ou mais métodos). Entretanto, a suavização exponencial foi utilizada como *benchmark*. E o MASE médio desta foi de 3,637, não muito distante do MASE das 40 séries da Tabela 1, cuja média é de 4,099 (12,7% superior).

Dessa forma, quando ocorrem replicações independentes, os resultados passam a ser válidos para um número maior de contextos. A consequência lógica é que eles devam ser válidos para um número maior de situações do que aquelas testadas até então. Assim, cada vez que uma hipótese não é rejeitada em testes especificamente projetados para refutá-la, ela se torna mais robusta ou com maior grau de corroboração (ALVES-MAZZOTI; GEWANDSZNAJDER 2002). Portanto, é provável que tais resultados sejam válidos para muitas séries encontradas na prática e que tenham características semelhantes às séries anuais da *M3e* da *M4 Competition*.

Uma diferença encontrada com relação *M4 Competition* é que nesta os métodos de RNA tiveram um desempenho inferior aos métodos estatísticos e ao *benchmark*. Porém Makridakis, Spiliotis e Assimakopoulos (2018) p. 805 explicam que os métodos de RNA utilizados eram apenas “diretrizes genéricas e algoritmos padrão que podem ser melhorados substancialmente”. Consequentemente, o presente artigo confirma esta afirmação e demonstra que o algoritmo utilizado pelo pacote *forecast* é eficiente.

Um resultado interessante do presente estudo é que, muito embora os três métodos tenham tido em média a mesma precisão, em algumas séries da tabela 1 (p. ex., série 11 ou 30), um dos métodos teve um desempenho muito diferente dos outros e se sobressaiu. Dessa forma, seria interessante em trabalhos futuros tentar verificar quais seriam as características distintivas dessas séries que levaram a tal resultado tão dispar e se há algum tipo de padrão.

Com relação às limitações do trabalho, a primeira delas é que os resultados são restritos a séries temporais sem sazonalidade. De acordo com Makridakis, Spiliotis e Assimakopoulos (2018), séries sazonais são mais fáceis de serem previstas, pois o efeito sazonal domina a flutuação da série, ao passo que em séries não sazonais, o principal fator são as variações irregulares (aleatórias). Assim, é possível que algum dos três métodos estudados neste trabalho levem alguma vantagem com relação aos demais em prever tal sazonalidade. Logo, uma possibilidade para trabalhos futuros seria investigar o desempenho dos métodos em séries com sazonalidade.

Outra limitação é que neste estudo foram utilizados métodos automáticos de seleção de modelos. É possível que um analista experiente utilizando informações a respeito das séries e os métodos tradicionais de identificação chegasse a modelos melhores do que os identificados automaticamente pelo *software* e que nesses modelos a precisão das previsões diferisse significativamente. Assim, os resultados obtidos devem ser interpretados neste contexto. Neste trabalho, o uso dos métodos automáticos de seleção se deveu à grande quantidade de séries analisadas simultaneamente. Dessa forma, uma possibilidade para trabalhos futuros seria avaliar o impacto do método de identificação (tradicional *versus* minimização de função penalizadora) na precisão das previsões.

Uma terceira limitação que enseja trabalhos futuros é relativa ao horizonte das previsões e o tipo de erro utilizado (MASE). Neste artigo, as comparações entre as precisões dos métodos foram baseadas no MASE de previsões 6 períodos a frente. Seria interessante verificar se os resultados aqui obtidos se mantêm variando o tipo de erro e os períodos de tempo das previsões, principalmente para previsões mais curtas.

Outra questão é que a diferença entre a precisão dos métodos possa depender de fatores característicos das séries tais como heteroscedasticidade, não linearidade e grau de aleatoriedade (MAKRIDAKIS; SPILIOTIS; ASSIMAKOPOULOS, 2018). Na amostra



utilizada, estes e demais fatores de ruído foram controlados/isolados usando o teste emparelhado. Assim, o resultado é prontamente generalizável para a população de 561 séries do IPEA e provavelmente válido para séries semelhantes. Um resultado muito mais abrangente poderia ser alcançado modelando-se explicitamente estes fatores utilizando a teoria estatística do planejamento de experimentos (MONTGOMERY; RUNGER, 2012). Esta última opção poderia ser alcançada em trabalhos futuros manipulando-se séries simuladas, ao invés de usar dados observacionais como no presente artigo.

Referências

- ALVES-MAZZOTTI, A. J.; GEWANDSZNAJDER, F. *O Método nas Ciências Naturais e Sociais: pesquisa quantitativa e qualitativa*. 2. Ed. São Paulo: Pioneira Thomson Learning, 2002.
- HYNDAMN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. 2. Ed. Melbourne: OTexts, 2018. Disponível em: <<https://otexts.org/fpp2/>>. Acesso em: 14 set. 2018.
- HYNDMAN R. J.; KHANDAKAR, Y. Automatic Time Series Forecasting: the forecast package for R. *Journal of Statistical Software*, 27(3), 2008.
- HYNDMAN, R. J.; KOEHLER, A. B.; ORD, J. K.; SNYDER, R. D. *Forecasting with exponential smoothing: the state space approach*. Berlin: Springer-Verlag, 2008.
- INSTITUTO DE PESQUISA ECONÔMICA APLICADA (IPEA). Ipeadata: disponível em <<http://www.ipeadata.gov.br>>. Acesso em: 4 fev. 2019.
- MAKRIDAKIS, S.; HIBON, M. The M3 Competition: results, conclusions and implications. *International Journal of Forecasting*. 16(4), p.451-476, 2000.
- MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. The M4 Competition: results, findings, conclusion and way forward. *International Journal of Forecasting* 34(4), p.802-808, 2018.
- MONTGOMERY, D. C.; RUNGER, G.C. *Estatística Aplicada e Probabilidade para Engenheiros*. 5. Ed. Rio de Janeiro: LTC, 2012.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. 2. Ed. São Paulo: Blucher, 2006.
- R CORE TEAM (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria.. Disponível em <<http://www.R-project.org>>, 2019.
- SIEGEL, S.; CASTELLAN JUNIOR, J. *Estatística não-paramétrica para ciências do comportamento*. 2. ed. Porto Alegre: Artmed, 2006.

Apêndice 1 – Comandos utilizados no software R para o cálculo dos erros de previsões

```
### Instala o pacote forecast (única vez)
install.packages("forecast", dependencies = TRUE)

### Carrega o pacote na memória no início de cada sessão
library(forecast)

### Pré-processamento dos dados
## A série x abaixo é a 20ª série analisada na Tabela 1
x = ts(c(124.24, 114.76, 105.65, 88.01, 75.56, 86.79, 94.64, 90.36,
123.13, 103.44, 100, 119.13, 114.62, 106.14, 101.45, 95.14, 96.29,
122.19, 123.38, 117.68, 67.91, 53.79), start = 1, frequency = 1)

n = length(x)
treino = window(x, start = 1, end = (n-6))
teste = window(x, start = (n-5), end = n)
```



Ajuste, previsão e erro de previsão

ARIMA

```
m1 = auto.arima(treino, stepwise = FALSE, approximation = FALSE)
p1 = forecast(m1, h = 6)
accuracy(p1, teste)
```

ETS

```
m2 = ets(treino)
p2 = forecast(m2, h = 6)
accuracy(p2, teste)
```

RNA

```
m3 = nnetar(treino)
p3 = forecast(m3, h = 6)
accuracy(p3, teste)
```